

## DISCRETIZACION (GRANULARIDAD) DEL CONOCIMIENTO

### Tesauros

- En el campo de la Bibliotecología o de las Ciencias de la Información, un tesauro es una lista de temas (en la forma de *términos*, es decir, pueden estar constituidos por más de una palabra), relacionados entre sí jerárquicamente (términos generales y subordinados), utilizadas para la indexación (con fines de archivo) y la recuperación de documentos.

En la época moderna, fue la Lexicografía el primer campo en el que se aplicó el término tesauro y en el que se elaboraron las primeras definiciones. En 1852 Peter Mark Roget publica su obra ROGET'S THESAURUS of English Words and Phrases (en lo sucesivo, tesauro ROGET'S ), cuya finalidad es la de proporcionar ayuda y asistencia a la expresión de ideas y a la composición literaria. El tesauro ROGET'S tiene una organización conceptual, donde a partir de los conceptos explicitados en las entradas se relacionan y agrupan las palabras que designan o nombran, en diversos contextos, al concepto en cuestión. Es a partir de esta conceptualización propuesta por Roget que comienzan a diversificarse las acepciones y usos del término tesauro.

En 1957 se usa por primera vez el término tesauro en el campo de la Biblioteconomía y Documentación, en un trabajo presentado a la Dorking Conference on Classification por Helen Brownson, miembro de la American National Science Foundation. En dicha contribución, el término tesauro es utilizado al analizar la problemática de traducir los conceptos y sus relaciones, tal como se expresan en los documentos, a un lenguaje con mayor precisión y sin ambigüedades, para facilitar la recuperación de información. Plantear una evolución histórica de la concepción de los tesauros como herramienta documental en el campo de la la Biblioteconomía y Documentación, excede a los propósitos del presente artículo, y, por lo tanto, nos limitaremos a seleccionar algunas definiciones que, desde nuestro punto de vista, permiten delinear sus principales características.

- En líneas generales, un tesauro comprende lo siguiente:
  - **Unidades léxicas:** que pueden comportar las siguientes categorías: grupos de descriptores que agrupan los términos de indexación bien por campos (temas), bien por clases de términos (facetas).
    - **descriptores:** términos permitidos. Un descriptor es una palabra o conjunto de palabras del lenguaje corriente que responden a un concepto. Esto es, los descriptores son términos o unidades lingüísticas que expresan conceptos. Un concepto se expresa con un único término y ese término responde a un único concepto. Dicho término será utilizado para el análisis y la recuperación del documento.
      - **términos simples o unitérminos:** se emplean cuando el concepto es claro en sí mismo, sin necesidad de añadir ninguna otra palabra, o cuando se usa como genérico de términos más concretos o específicos. Ejemplo: **Documentos**.
      - **términos compuestos:** la mayoría de los términos de un tesauro adquieren su verdadero sentido cuando se convierten en términos compuestos, ya que así se logra una mayor especificidad. La estructura sintáctica de los términos compuestos suele ser la siguiente:
        - sustantivo + adjetivo: **Documentos digitales**
        - sustantivo + sintagma preposicional: **Documentos de archivo**
        - sustantivo + sintagma aposicional: **Documentos RDF/XML**
    - **no descriptores:** términos prohibidos. Al elegir un único término para eliminar la sinonimia del lenguaje natural, se rechazan los términos restantes, que serán los no descriptores.
  - **Relaciones semánticas:** son todas aquellas relaciones que se establecen entre los términos de un tesauro. Podemos distinguir las siguientes:
    - **relaciones de equivalencia:** se derivan de la gran proximidad fonética o sémica entre los descriptores y no descriptores (sinónimos, cuasi-sinónimos, etc). Este tipo de relaciones se suelen representar mediante el término USE (relación entre el término no preferente y el término preferente, o término a usar más aceptado) y las siglas UP (Usado por),

que en inglés son USE y UF. Ejemplo: Dictablanda **GOBIERNO BERENGUER**. USE **GOBIERNO BERENGUER** U.P (Usado por) Dictablanda

- **relaciones jerárquicas:** son relaciones asimétricas, establecidas entre los descriptores, en la que un término es superior o genérico de otro. Esta estructura clasificatoria distingue al tesoro de las listas alfabéticas de materias o palabras clave, así como de diccionarios y léxicos. De esta forma se crean campos conceptuales que clasifican los términos y los agrupan en un campo delimitado de significados o facetas que se van agrupando en disciplinas y sub-disciplinas. Para la presentación de las relaciones jerárquicas se suelen utilizar determinadas siglas como T.G. (Término Genérico) y T.E. (Término Específico); o por su siglas en inglés, BT (Broad Term o término más amplio), NT (Narrow Term o término más específico), SY (sinónimo), RT (término relacionado), etc. Si un descriptor tiene un sólo descriptor por encima de él es monojerárquico, si tiene dos o más es polijerárquico. Las relaciones jerárquicas pueden ser de dos tipos:
  - **Relaciones género-especie:** se aplica a acciones, propiedades y agentes. (T.G. Armas/T.E. Armas blancas)
  - **Relaciones parte-todo:** el nombre de una parte implica y evoca la del todo poseedor. Se aplica a objetos, lugares, disciplinas y sus partes de estudio, estructuras sociales, etc. (T.G. Guerra/T.E. Guerra de Irak)
- **relaciones asociativas:** son las relaciones asimétricas establecidas entre descriptores de distintas familias, distintas a las jerárquicas o de equivalencia, y que designan conceptos capaces de evocar mutuamente, por asociación de ideas, relaciones entre ellos. Este tipo de relación se suele representar mediante las siglas T.R. (Término Relacionado) o en inglés RT (Related Term). La mayoría de las relaciones asociativas se establecen entre términos que comparten un lexema común (por ejemplo, POLÍTICA. T.R. POLÍTICOS) o que mantienen unas relaciones de hiperonimia o hiponimia (CARGOS MUNICIPALES. T.R. ADMINISTRACIÓN LOCAL). También podemos añadir Notas de Alcance (N.A.) con el fin de ofrecer una explicación o descripción del significado del término.

- El tesoro es un recurso alternativo del árbol de conocimiento
- Es un símil del diccionario de datos, cuando se piensa en una base de conocimientos informatizada
- Tanto el tesoro como el árbol pueden incorporar información sobre interrelaciones semánticas entre los conceptos, estadísticas de la frecuencia de uso, información asociada a la ubicación geográfica, etcétera.

Después de lo señalado anteriormente podemos concluir que:

Un tesoro es "un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento. El tesoro proporciona una organización semántica principalmente a través de la explicitación de las relaciones establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los términos que los representan. Los tesoros, para la Documentación y la Biblioteconomía, son por lo tanto una herramienta de control terminológico, puesto que la estructuración del tesoro es en base a conceptos. Los conceptos, son representados por términos seleccionados que evidencian ese control terminológico. Con dicho control se busca neutralizar la sinonimia y la polisemia, ambas características naturales de la lengua, que dificultan la precisión de la indización y recuperación de la información, ambas funciones básicas de los tesoros. Los tesoros son herramientas construidas para el auxilio tanto de los profesionales de la información como de los usuarios finales".

### **Los tesoros y el entorno digital**

Como se ha mencionado en el apartado anterior, el impacto tecnológico del entorno digital ha afectado por igual a recursos, herramientas y comportamiento del usuario en relación a la representación y recuperación de información. Esta influencia abre nuevas posibilidades en cuanto a las pautas de diseño, elaboración, gestión y uso de las herramientas utilizadas para la representación y recuperación de información. Desde esta perspectiva se han empezado a perfilar los aportes del entorno digital a los tesoros, que podemos resumir en los siguientes elementos.

- Un primer elemento a tener en cuenta es el enriquecimiento de la funcionalidad de la estructura de los tesauros a partir de la hipertextualidad. Esto se traduce en el establecimiento de hipervínculos, entre todos los elementos estructurales (descriptores, no descriptores, notas de alcance, etc.), y también entre las distintas partes del tesoro.
- Un segundo elemento es la reducción de costos de actualización y mantenimiento. Debido a la creciente informatización de los procesos de construcción de tesauros y al progresivo abandono del soporte papel para la publicación de dichas herramientas, es viable una reducción de costos.
- Un tercer elemento, es la integración del usuario al proceso de creación, gestión y optimización de los tesauros, a través de tests de usabilidad, uso de técnicas de modelado de usuario, etc. Esto permite elaborar herramientas que tengan en cuenta los requerimientos de los usuarios, y descartar su construcción como simples estructuras teóricas.
- Un cuarto elemento es la posibilidad de aplicar medidas de reutilización e interoperabilidad en el momento de planificar y construir los tesauros. Con ello se posibilita el aprovechamiento y enriquecimiento de la información conceptual y lingüística ya generada para otros recursos.

### ***La ontología***

La Filosofía es el primer campo de conocimiento donde se utiliza el concepto de ontología. El uso de dicho concepto se origina en la noción aristotélica de filosofía primera -luego llamada metafísica-, que por un lado estudia la esencia del ser ( el ser como ser) , y por otro lado, estudia las características básicas de toda la realidad como un todo ( el ser o el ente principal del cual dependen los demás entes). En la actualidad, es considerada una rama de la Filosofía que tiene por objetivo la explicación de la existencia de una manera sistemática, trata sobre los tipos y estructuras de objetos, propiedades, eventos, procesos y relaciones relativas a cada porción de la realidad.

En un sentido muy general, las ontologías son para la Inteligencia Artificial recursos construidos que permiten representar el conocimiento compartido y común sobre algo. Esta posibilidad de generar recursos compatibles y la consecuencia natural de intercambiar la información almacenada, es lo que provoca que un concepto como el de ontología (antes privativo de la Inteligencia Artificial) se filtre en los entornos de trabajo de otros ámbitos y en especial en lo relativo a la gestión de los recursos y herramientas del entorno digital y sobre todo a la gestión de Objetos de Aprendizaje.

### ***Relación entre los conceptos de tesoro y ontología***

En las secciones anteriores se mencionaron las características de los tesauros y ontologías, y hemos propuesto una conceptualización para cada una de estas herramientas de representación y recuperación de información. Del análisis realizado de la selección de definiciones, concluimos que:

- Un tesoro es una herramienta documental utilizada en el ámbito de la representación y recuperación de información, que representa un ámbito del conocimiento determinado mediante su estructuración conceptual. La estructura conceptual proporciona una organización semántica mediante la explicitación de las relaciones conceptuales y de la restricción del significado de los términos que los representan. El ámbito de conocimiento se estructura mediante relaciones conceptuales de tipo jerárquico, asociativo y de equivalencia. Un tesoro es utilizado tanto por usuarios profesionales de la información como por usuarios finales.
- Una ontología es una representación formal y explícita de la estructura conceptual de un campo del conocimiento. Una ontología es el soporte semántico de las palabras que son descritas como objetos lingüísticos en una base de datos léxica o terminológica. Las relaciones conceptuales que son representadas en una ontología son muy variadas y dependientes del campo del conocimiento a estructurar. Una ontología se construye a fin de compartir y reutilizar la información almacenada, la cual, por estar formalizada puede ser interpretada tanto por personas como por programas informáticos.

La ontología aporta al proceso de construcción de herramientas documentales para la representación y la recuperación de información, la declaración explícita de las relaciones

conceptuales en un ámbito en particular a través de la formalización semántica de dicha estructura.

La formalización semántica, a su vez, posibilita una representación lógica y coherente de dicha estructura conceptual, por lo tanto permite una interpretación informática que se orienta hacia la interoperabilidad y reutilización por parte de otros recursos o aplicaciones. La formalización semántica, por lo tanto, implica traducir a una sintaxis legible por ordenador el enunciado que aporte el significado del tipo de relación existente entre dos o más conceptos.

## **CARTOGRAFÍA CONCEPTUAL**

- Una cartografía conceptual es el conjunto de mapas conceptuales que determinan el conjunto de dominios del conocimiento de una disciplina profesional o de un campo científico o cultural
- Podemos establecer que un plan de estudios tiene asociada una cartografía conceptual

### ***Mapas Conceptuales***

Los mapas conceptuales son herramientas de especificación y visualización de dominios del conocimiento, de gran riqueza

- Son equivalentes a los árboles de conocimiento
- Los árboles representan la estructura de datos con que se almacenan los objetos
- Esta estructura es la base para almacenar, buscar y recuperar los objetos